His-HA tagged Def1 (Def1-decahistidine-HA, Def1HH) had wild-type phenotypes. For purification purposes, tagging was done in the protease-deficient PY26 strain. Details of affinity tagging are available on request.

To construct *def1 rpo21* strains, *DEF1* was replaced with the *LEU2* marker in YF2277 (ref. 9) (containing the wild-type *RPO21* gene on a *URA3 CEN* plasmid). These cells and control YF2277 cells were transformed with *TRP1 CEN* plasmids expressing *rpo21* alleles[9]. *ura3*$^-$ cells were selected on 5-FOA-containing media.

Protein purification and identification

During purification, Def1HH was detected by 12CA5 (anti-HA) antibody, whereas MHRad26 was detected either by 9E10 (anti-Myc) antibody or by a polyclonal Rad26 antibody. Yeast extract preparations and chromatography of proteins on resins such as BioRex-70 (Bio-Rad), nickel-agarose (Qiagen), DEAE Sepharose (Pharmacia), and MonoQ HR5/5 (Pharmacia) was done as previously described for other factors[5]. Approximately 1 kg of yeast paste from the strains indicated above was processed. Rad26 from the soluble (DNA-free) fraction was eluted in the 600 mM step from BioRex-70. After nickel-agarose chromatography, eluted proteins were bound to 9E10-adsorbed protein A-Sepharose (Pharmacia), washed with A-500 (ref. 5), and eluted with either 150 mM NaCl, 100 mM glycine pH 2.5; or by tobacco etch virus (TEV) protease digestion according to the manufacturers' instructions (Gibco Life Technologies). Def1HH from the soluble fraction was recovered in BioRex-70 flow-through and was loaded directly onto a DEAE Sepharose resin. After washing with B-150 (ref. 5), proteins were eluted with B-1000 (lacking dithiothreitol (DTT) and EDTA), and incubated with nickel-agarose overnight. After elution with B-500 (lacking DTT and EDTA) containing 200 mM imidazole, the eluate was adsorbed to 12CA5-adsorbed protein A-Sepharose, washed with A-300, and eluted as described[6].

Purification from the chromatin fractions was done as follows: release of proteins from salt-stable (500 mM KOAc) soluble chromatin fragments by treatment with 1 M ammonium sulphate (and in some cases also RNAse and DNAse treatment) was done as described[5]. Purification on nickel-agarose and antibody-affinity resins was done as described above. Proteins eluted from 12CA5-affinity resin (Def1HH purification) were further fractionated by loading onto MonoQ HR5/5, which was resolved by a ten-column volume gradient from 100 to 1,000 mM KOAc in buffer B (ref. 5).

Other techniques

Protein identification was done after trypsin digestion of gel-fractionated proteins as described[6]. A peptide representing the C-terminal 20 amino acids of Rad26 was cross-linked to keyhole limpet haemocyanin (Calbiochem) and used to immunize rabbits (Murex) for antibody production. TCR assays were performed as described[29]. Methods for investigating RNAPII degradation were as described[12], except that UV-irradiation ($30\,\mathrm{J\,m^{-2}}$) was done on cells re-suspended in saline. Detection of ubiquitinated RNAPII by western blotting was done after immunoprecipitation of polymerase using 8WG16 (ref. 30) antibodies.

Received 10 October; accepted 12 December 2001.

1. Friedberg, E. C., Walker, G. C. & Siede, W. *DNA Repair and Mutagenesis* (ASM Press, Washington, DC, 1995).
2. Hanawalt, P. C. Controlling the efficiency of excision repair. *Mutat. Res.* **485,** 3–13 (2001).
3. de Boer, J. & Hoeijmakers, J. H. Nucleotide excision repair and human syndromes. *Carcinogenesis* **21,** 453–460 (2000).
4. van Gool, A J. *et al.* RAD26, the functional *S. cerevisiae* homolog of the Cockayne syndrome B gene ERCC6. *EMBO J.* **13,** 5361–5369 (1994).
5. Otero, G. *et al.* Elongator, a multi-subunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. *Mol. Cell* **3,** 109–118 (1999).
6. Winkler, G. S. *et al.* RNA polymerase II elongator holoenzyme is composed of two discrete subcomplexes. *J. Biol. Chem.* **276,** 32743–32749 (2001).
7. Verhage, R. A. *et al.* Double mutants of *Saccharomyces cerevisiae* with alterations in global genome and transcription-coupled repair. *Mol. Cell Biol.* **16,** 496–502 (1996).
8. Wind, M. & Reines, D. Transcription elongation factor SII. *Bioessays* **22,** 327–336 (2000).
9. Archambault, J., Lacroute, F., Ruet, A. & Friesen, J. D. Genetic interaction between transcription elongation factor TFIIS and RNA polymerase II. *Mol. Cell Biol.* **12,** 4142–4152 (1992).
10. Wu, J., Awrey, D. E., Edwards, A. M., Archambault, J. & Friesen, J. D. *In vitro* characterization of mutant yeast RNA polymerase II with reduced binding for elongation factor TFIIS. *Proc. Natl Acad. Sci. USA* **93,** 11552–11557 (1996).
11. Ratner, J. N., Balasubramanian, B., Corden, J., Warren, S. L. & Bregman, D. B. Ultraviolet radiation-induced ubiquitination and proteasomal degradation of the large subunit of RNA polymerase II. Implications for transcription-coupled DNA repair. *J. Biol. Chem.* **273,** 5184–5189 (1998).
12. Beaudenon, S. L., Huacani, M. R., Wang, G., McDonnell, D. P. & Huibregtse, J. M. Rsp5 ubiquitin-protein ligase mediates DNA damage-induced degradation of the large subunit of RNA polymerase II in *Saccharomyces cerevisiae. Mol. Cell Biol.* **19,** 6972–6979 (1999).
13. Lommel, L., Bucheli, M. E. & Sweder, K. S. Transcription-coupled repair in yeast is independent from ubiquitylation of RNA pol II: implications for Cockayne's syndrome. *Proc. Natl Acad. Sci. USA* **97,** 9088–9092 (2000).
14. Conaway, J. W., Shilatifard, A., Dvir, A. & Conaway, R. C. Control of elongation by RNA polymerase II. *Trends Biochem. Sci.* **25,** 375–380 (2000).
15. Jelinsky, S. A., Estep, P., Church, G. M. & Samson, L. D. Regulatory networks revealed by transcription profiling of damaged *Saccharomyces cerevisiae* cells: Rpn4 links base excision repair with proteasomes. *Mol. Cell Biol.* **20,** 8157–8167 (2000).
16. Hilt, W. & Wolf, D. H. Proteasomes: destruction as a programme. *Trends Biochem. Sci.* **21,** 96–102 (1996).
17. Russell, S. J., Reed, S. H., Huang, W., Friedberg, E. C. & Johnston, S. A. The 19S regulatory complex of the proteasome functions independently of proteolysis in nucleotide excision repair. *Mol. Cell* **3,** 687–695 (1999).
18. Gillette, T. G. *et al.* The 19S complex of the proteasome regulates nucleotide excision repair in yeast. *Genes Dev.* **15,** 1528–1539 (2001).
19. Ferdous, A., Gonzalez, F., Sun, L., Kodadek, T. & Johnston, S. A. The 19S regulatory particle of the proteasome is required for efficient transcription elongation by RNA polymerase II. *Mol. Cell* **7,** 981–991 (2001).
20. Bregman, D. B. *et al.* UV-induced ubiquitination of RNA polymerase II: a novel modification deficient in Cockayne syndrome cells. *Proc. Natl Acad. Sci. USA* **93,** 11586–11590 (1996).
21. McKay, B. C. *et al.* UV light-induced degradation of RNA polymerase II is dependent on the Cockayne's syndrome A and B proteins but not p53 or MLH1. *Mutat. Res.* **485,** 93–105 (2001).
22. Luo, Z., Zheng, J., Lu, Y. & Bregman, D. B. Ultraviolet radiation alters the phosphorylation of RNA polymerase II large subunit and accelerates its proteasome-dependent degradation. *Mutat. Res.* **486,** 259–274 (2001).
23. Schmickel, R. D., Chu, E. H., Trosko, J. E. & Chang, C. C. Cockayne syndrome: a cellular sensitivity to ultraviolet light. *Pediatrics* **60,** 135–139 (1977).
24. Venema, J., Mullenders, L. H., Natarajan, A. T., van Zeeland, A. A. & Mayne, L. V. The genetic defect in Cockayne syndrome is associated with a defect in repair of UV-induced DNA damage in transcriptionally active DNA. *Proc. Natl Acad. Sci. USA* **87,** 4707–4711 (1990).
25. Mayne, L. V. & Lehmann, A. R. Failure of RNA synthesis to recover after UV irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma pigmentosum. *Cancer Res.* **42,** 1473–1478 (1982).
26. van Gool, A. J. *et al.* The Cockayne syndrome B protein, involved in transcription-coupled DNA repair, resides in an RNA polymerase II-containing complex. *EMBO J.* **16,** 5955–5965 (1997).
27. Yu, A., Fan, H. Y., Liao, D., Bailey, A. D. & Weiner, A. M. Activation of p53 or loss of the Cockayne syndrome group B repair protein causes metaphase fragility of human U1, U2, and 5S genes. *Mol. Cell* **5,** 801–810 (2000).
28. Le Page, F. *et al.* Transcription-coupled repair of 8-oxoguanine: requirement for XPG, TFIIH, and CSB and implications for Cockayne syndrome. *Cell* **101,** 159–171 (2000).
29. Jansen, L. E. *et al.* Spt4 modulates Rad26 requirement in transcription-coupled nucleotide excision repair. *EMBO J.* **19,** 6498–6507 (2000).
30. Thompson, N. E. & Burgess, R. R. Immunoaffinity purification of RNA polymerase II and transcription factors using polyol-responsive monoclonal antibodies. *Methods Enzymol.* **274,** 513–526 (1996).

Acknowledgements

We thank the Cancer Research UK service facilities, especially fermentation services, for their help. We thank A. Grewal for assistance with peptide mass fingerprinting, and B. Winkler, P. Verrijzer, S. West and T. Lindahl for comments on the manuscript. This work was supported by grants from the ICRF and the Human Frontier Science Programme to J.Q.S., and by a NCI Core Grant to P.T.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.Q.S. (e-mail: j.svejstrup@cancer.org.uk).

Structural basis for acidic-cluster-dileucine sorting-signal recognition by VHS domains

Saurav Misra*, Rosa Puertollano†, Yukio Kato†, Juan S. Bonifacino† & James H. Hurley*

* *Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA*

† *Cell Biology and Metabolism Branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA*

Specific sorting signals direct transmembrane proteins to the compartments of the endosomal–lysosomal system[1]. Acidic-cluster-dileucine signals present within the cytoplasmic tails of sorting receptors, such as the cation-independent and cation-dependent mannose-6-phosphate receptors, are recognized by the GGA (Golgi-localized, γ-ear-containing, ADP-ribosylation-factor-binding) proteins[2–5]. The VHS (Vps27p, Hrs and STAM) domains[6] of the GGA proteins are responsible for the highly specific recognition of these acidic-cluster-dileucine signals[7–10]. Here we report the structures of the VHS domain of human GGA3

complexed with signals from both mannose-6-phosphate receptors. The signals bind in an extended conformation to helices 6 and 8 of the VHS domain. The structures highlight an Asp residue separated by two residues from a dileucine sequence as critical recognition elements. The side chains of the Asp-X-X-Leu-Leu sequence interact with subsites consisting of one electropositive and two shallow hydrophobic pockets, respectively. The rigid spatial alignment of the three binding subsites leads to high specificity.

We expressed the VHS domain from human GGA3 and crystallized it with signal peptides derived from the carboxy termini of the cation-independent mannose-6-phosphate receptor (CI-MPR; FHDDSDEDLLHI) and the cation-dependent mannose-6-phosphate receptor (CD-MPR; EESEERDDHLLPM; Fig. 1a). These crystals diffracted to 2.4 and 2.2 Å, respectively, permitting the identification of the binding sites for the key residues of the acidic-cluster-dileucine sorting sequences on the VHS domain. In each structure there are four protein molecules in the asymmetric unit, which provide a view of the interactions of each peptide in four slightly different crystal packing environments. In spite of the tetramer observed in the asymmetric unit, the GGA3 VHS domain is a monomer in solution, as judged by analytical ultracentrifugation at physiological pH and ionic strength in the presence or absence of the CI-MPR peptide (R. Ghirlando, personal communication), and we believe that it functions as a monomer in peptide binding.

Like the VHS domains of Hrs[11] and Tom1[12] (Target of Myb1), the GGA3 VHS domain is a right-handed superhelix of eight helices (Fig. 1b). The structure differs from those of the Hrs and Tom1 VHS domains by 1.5-Å root-mean-square deviation (Cα positions). Both the CI-MPR and CD-MPR peptides bind in an extended conformation to a surface formed by helices 6 and 8 (Fig. 1b–e). These helices are approximately parallel to each other, touching at their amino termini but diverging at their C termini far enough for a groove to form between them. We consider the most buried acidic residue of the MPR peptides to be in consensus position 0, with more N-terminal positions indicated by negative numbers. Positions –6 to –3 are disordered. The remaining residues are in an extended conformation and make extensive contacts with helices 6 and 8. The first well-ordered N-terminal parts of the signal peptides (residues –2 to 0) bind near the N terminus of helix 6. The central parts of the peptides (residues 1 and 2) extend over the point of contact between helices 6 and 8. Finally, the C-terminal parts (residues 3–6) drop into the groove formed between the C-terminal ends of the two helices.

The acidic cluster in the N-terminal part of the motif is the first of the motif's two defining features. Of the acidic residues at positions –2 to 2, Asp 0 forms the most extensive interactions (Fig. 2a, b). Its side chain interacts with the N terminus of helix 6, making hydrogen bonds with the backbone amides of GGA3 residues Phe 87 and Arg 88, and a salt bridge with Lys 130. The extensive interactions are consistent with the crucial role for this Asp residue in MPR sorting, as judged from mutagenesis[13]. Asp 0 is in a sterically restricted environment, which explains why even replacement with a Glu residue is not tolerated at this position[13]. Nevertheless, the Asp 0 side chain is only partly buried on binding, losing 83% and 71% of its

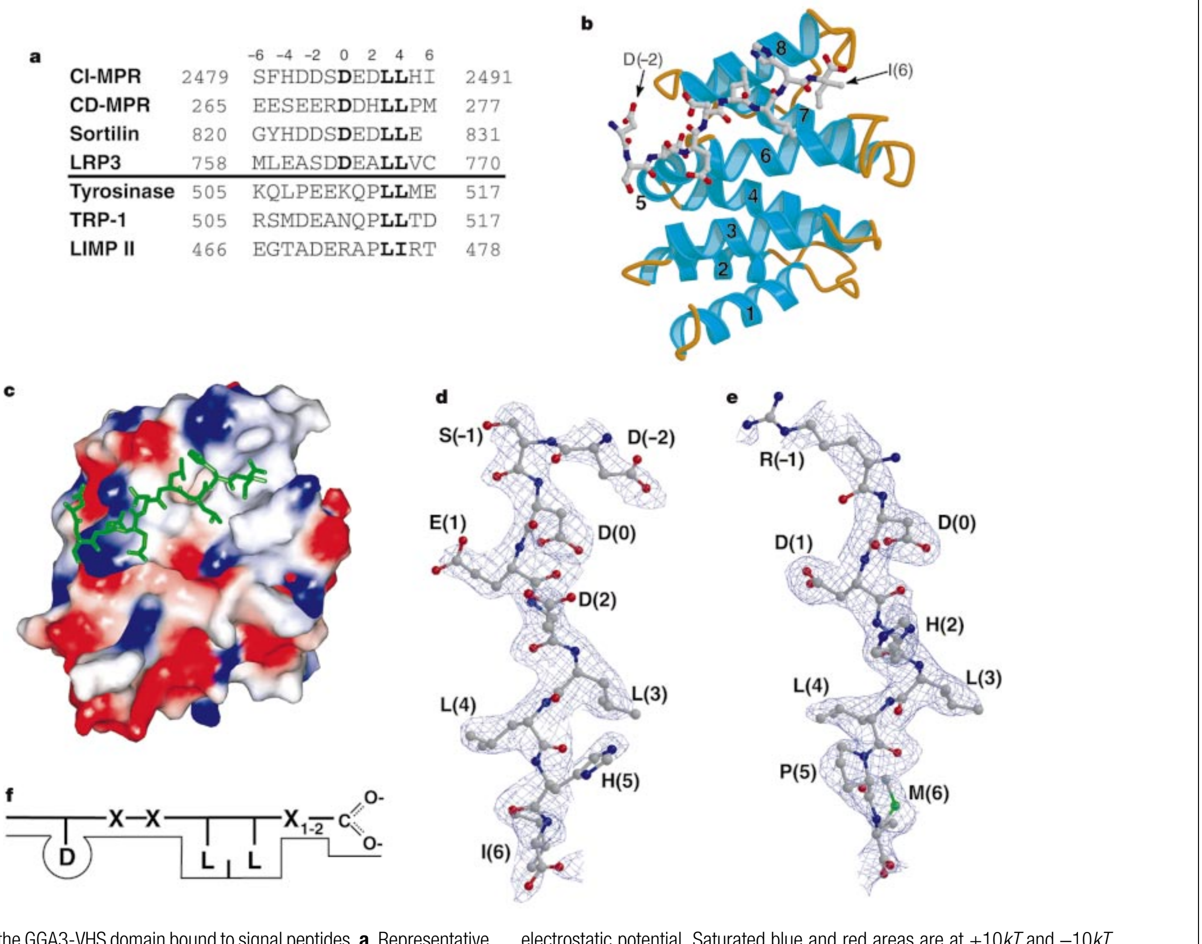

Figure 1 Structures of the GGA3-VHS domain bound to signal peptides. **a**, Representative acidic-cluster-dileucine motifs. The first four sequences bind GGA-VHS domains[7–10], the last three do not[7]. **b**, Structure of the VHS domain with CI-MPR sorting signal (ball-and-stick model) bound between helices 6 and 8. The first and last visible residues of the CI-MPR peptide are labelled. **c**, Molecular surface of the VHS domain, coloured by electrostatic potential. Saturated blue and red areas are at $+10kT$ and $-10kT$ respectively. The CI-MPR peptide is shown in green. **d**, **e**, Calculated $2f_o - f_c$ omit maps for the CI-MPR (**d**) and CD-MPR (**e**) sorting signals, contoured at 1.0σ and 0.9σ, respectively. **f**, Scheme highlighting the primary determinants of peptide binding to the GGA-VHS domains.

solvent-accessible surface area in the CI-MPR and CD-MPR peptide complexes, respectively. The interactions of other acidic residues with the VHS domain are weaker. Asp −2 is clearly resolved only in the peptide derived from CI-MPR and only in one of the four molecules in the asymmetric unit, where it interacts with the side chain of Lys 130. The acidic residue at position 1 is Glu in the CI-MPR peptide; in two of the four molecules in the asymmetric unit it interacts with the Nδ of Asn 91. In the CD-MPR peptide this residue is Asp and does not form these interactions. The side chain of Asp 2 in the CI-MPR peptide is orientated away from the VHS domain.

The two leucine residues of the motif are required for binding to the GGA VHS domains and for proper sorting[7–10,13,14]. The leucine residues are largely buried, with a loss of solvent-accessible surface area ranging from 72% to 77% for the two Leu residues in the two complexes. Leu 3 and Leu 4 bind in two adjacent shallow hydrophobic pockets formed by side chains from helices 6 and 8. Leu 3 interacts primarily with the side chains of Phe 87 and Ile 94 on helix 6, and Ala 134 and Met 137 on helix 8. Leu 4 interacts exclusively with residues on helix 6, including Tyr 101, Ile 94 and the aliphatic portions of the Asn 91 and Lys 95 side chains. The incomplete burial of the key Leu side chains is consistent with the finding that replacement with Ile decreases, but does not abolish, function[13].

GGA VHS domains interact with signals positioned very close to the C terminus of the protein, with the C-terminal residue typically

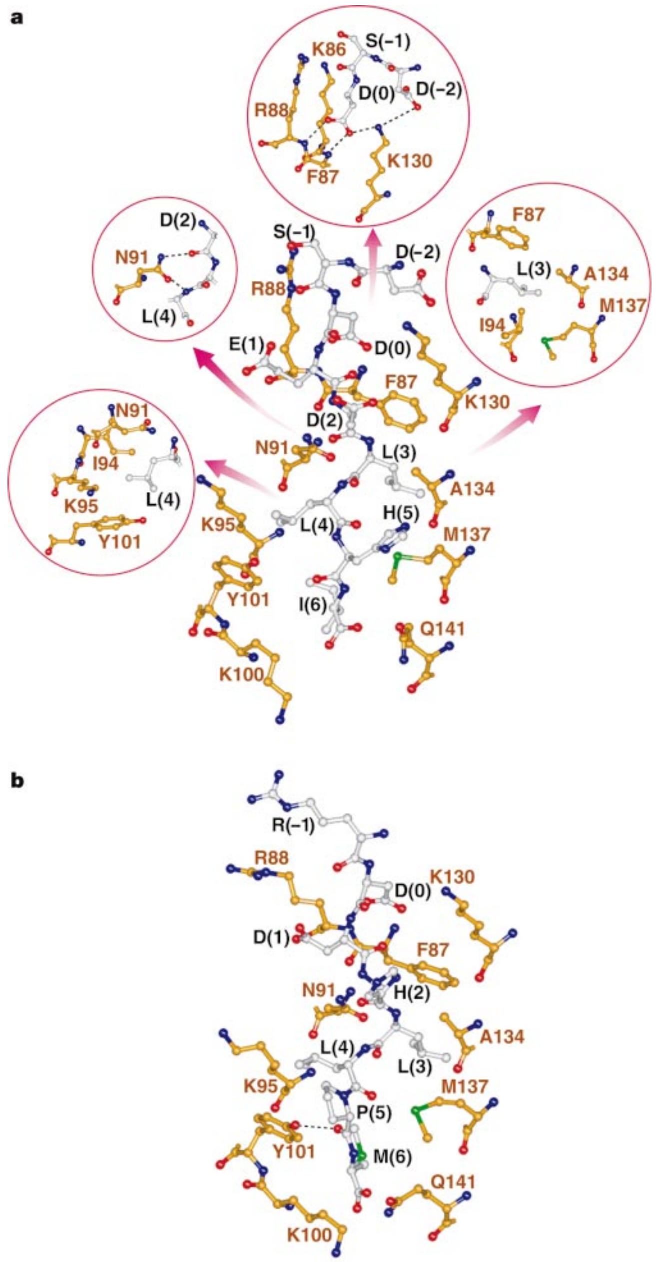

Figure 2 Molecular details of the interactions between the acidic-cluster-dileucine motifs and their binding site on the VHS domain. **a**, **b**, Ball-and-stick representations of the CI-MPR (**a**) and CD-MPR (**b**) sorting sequences, and the VHS residues with which they interact. Oxygen, nitrogen and sulphur atoms are coloured red, blue and green, respectively. Carbon atoms of the signal sequences are grey, and carbon atoms from VHS domain residues are coloured gold. Hydrogen bonds and salt-bridge interactions are shown as dashed lines. The arrows designate close-up views of the appropriate sites. Close-ups are not shown in **b** because the sites are very similar to the corresponding sites in **a**.

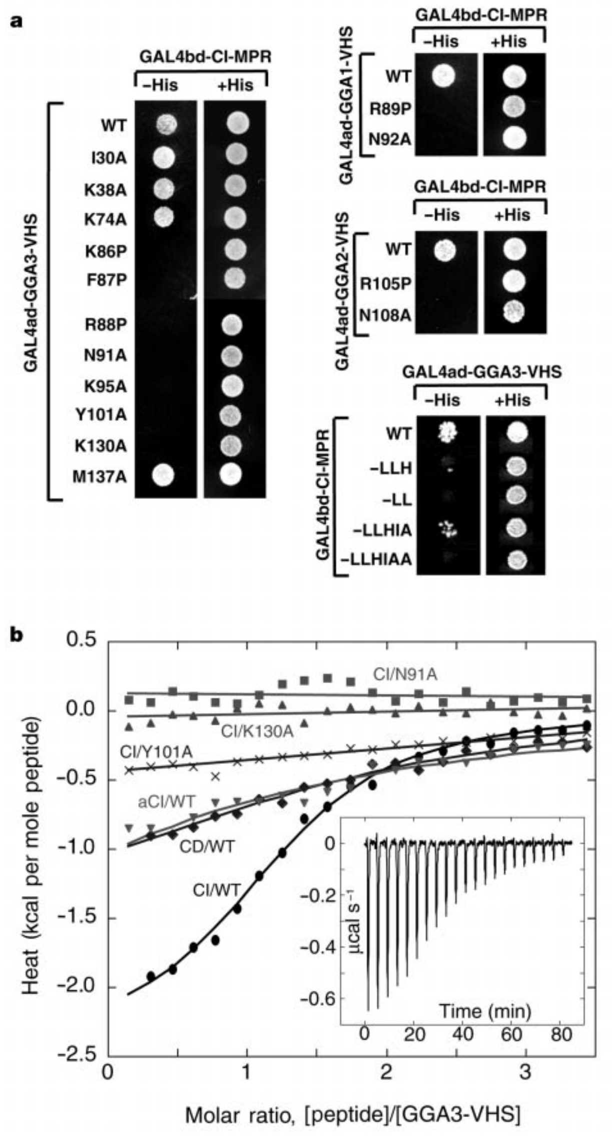

Figure 3 Functional analysis of interactions between the CI-MPR tail and GGA-VHS domains. **a**, Yeast two-hybrid experiments showing interactions between the CI-MPR tail and mutants of the GGA-VHS domains. Growth on –His plates is indicative of an interaction. GGA1 Arg 89/Asn 92 and GGA2 Arg 105/ Asn 108 are equivalent to GGA3 Arg 88/Asn 91 (see Supplementary Information). The bottom right panel depicts interactions between mutant CI-MPR tail constructs and wild-type (WT) GGA3-VHS. The C-terminal sequence of each tail construct is shown. **b**, Isothermal titration calorimetry measuring the binding of peptides to wild-type and mutant GGA3 VHS domains. aCI, C-terminally amidated CI-MPR peptide. Binding constants are listed in Table 1. The inset shows the differential heat released when CI-MPR peptide (1 mM) was injected into the GGA3-VHS solution (50 μM) in 10-μl aliquots. Trace shown after subtraction of data from the injection of peptide into a buffer blank.

at position 5 or 6. Residues 5 and 6 of the CD-MPR peptide are visible in only two out of four molecules of the asymmetric unit in our structure. The C termini of the peptides are close to several polar side chains adjoining a hydrophobic pocket near the C termini of helices 6 and 8. The C-terminal carboxyl group is orientated towards the side chains of Lys 100 and Gln 141, but not close enough to hydrogen bond with either of them. A hydrogen bond between the hydroxy group of Tyr 101 and the main-chain carbonyl group of residue 5 is formed with the C-terminal portion of the CD-MPR peptide but not with the CI-MPR peptide. The C-terminal side chain (Ile in CI-MPR; Met in CD-MPR) protrudes into a hydrophobic pocket bounded by the Ile 94, Ser 98 and Tyr 101 side chains. MPR sequences with two flanking residues are optimal; those with one or three bind more weakly, and those with zero or four do not bind (Fig. 3a). The identities of the flanking residues are not crucial[7,13]. C-terminally amidated variants of the CI-MPR and CD-MPR peptides bind very weakly as assessed by isothermal titration calorimetry (ITC) (Fig. 3b). Thus, the spacing between the LL sequence and the free C terminus is more important than the identities of the C-terminal side chains.

Position −1 has Ser in the CI-MPR peptide and Arg in the CD-MPR peptide. Ser −1 of CI-MPR is phosphorylated *in vivo*[15]. The Ser Oγ is 3.9 Å from the nearby basic residue Arg 88, too far for a hydrogen bond. However, the gap between the Ser and the Arg residues allows room for binding covalently attached phosphate. Phosphorylation of this Ser residue could position the resulting phosphoserine to interact strongly with Arg 88 and also Lys 86, indicating a mechanism for regulating the binding affinity of CI-MPR to the GGA domains.

Extended conformations of protein-bound peptides are frequently stabilized by β-sheet interactions between the peptide and an unpartnered β strand from the protein, as occurs in the binding of YXXØ (where Ø indicates a bulky hydrophobic residue) signals to the μ2 subunit of AP-2 (ref. 16). In the all-helical VHS domain, the extended conformation of the signal peptides is stabilized by Asn 91, which undergoes hydrogen-bonding to the backbone carbonyl group of residue 2 and the backbone amide group of Leu 4. Similar Asn–backbone interactions are seen in other helical protein–peptide complexes. The peroxisomal targeting signal PTS1 binds to a groove along two parallel helices of a helical tetratricopeptide repeat of its receptor PEX5 (ref. 17). The PTS1 backbone interacts with four Asn residues. A similar binding mode is observed for binding of the C termini of the heat-shock proteins Hsp70 and Hsp90 to three parallel helices of tetratricopeptide repeats of their adaptor protein, Hop[18].

To assess the functional relevance of the peptide-binding site observed in our structures, we examined the interaction between the CI-MPR tail and mutants of the GGA1-3 VHS domain by using the yeast two-hybrid system (Fig. 3a) We used ITC (see Fig. 3b and Table 1) to measure the binding of GGA1 and GGA3 to peptides, but the GGA2 VHS domain was poorly soluble and less tractable for ITC studies. The mutation to alanine of several residues that interact with key signal residues 0, 3 and 4 via their side chains (namely Asn 91, Lys 95, Tyr 101 or Lys 130) all abrogated binding. Residues involved in backbone interactions with the signal (namely Phe 87 or Arg 88) were mutated to proline, also abolishing binding. Substitution of Met 137, which is partly buried in hydrophobic contacts with Leu 3, did not abolish binding. As a control, we demonstrated that the mutation to alanine of residues that are not part of the signal-binding site (namely Ile 30, Lys 38 or Lys 74) had no effect on the interactions. As found also by surface plasmon resonance[7], the affinity of the signals for the VHS domains measured by ITC is in the 10-μM range (Table 1). In all cases the stoichiometry of binding was 1.2 ± 0.1, in agreement with the 1:1 binding seen in the crystals.

The GGA3 VHS residues that bind the MPR signals are completely conserved among the three human GGAs and are mostly conserved in *Drosophila melanogaster* and *Caenorhabditis elegans* GGA homologues (Supplementary Information). The structural inference that interactions observed in this structure also exist in the relevant complexes with GGA1 and GGA2 is borne out by mutational and binding analysis (Fig. 3) and three-dimensional modelling (data not shown) of the human GGA1 and GGA2 counterparts of the key residues Arg 88 and Asn 91. Only some of these residues are conserved in yeast Gga1p and Gga2p, indicating that the yeast proteins might bind motifs other than those described here. These residues are poorly conserved or not conserved at all in the remainder of the VHS domains. These include TOM1, TOM1-like, Hrs and STAM, which do not bind the MPR or sortilin tails[7,9]. The biological roles of VHS domains other than those of the GGAs remain unknown.

The structure focuses attention on Asp 0, Leu 3 and Leu 4 of the acidic-cluster-dileucine motif as the primary contributors to recognition. The strict requirement for Asp 0 in the signals explains why other dileucine-based signals, such as those found in the cytosolic tails of tyrosinase, TRP-1 (tyrosinase-related protein-1) and LIMP II (lysosomal integral membrane protein II) (Fig. 1a, bottom three sequences), fail to bind to the VHS domains of the GGAs[7]. These signals lack the critical Asp residue at position 0 but instead have Asp or Glu residues at positions −2 or −1, which are less important for interactions with the GGA VHS domains. The residues at position −1, −2 and 1 and the C-terminal carboxyl group contribute to binding in a secondary capacity. A critical GGA Asn residue stabilizes the extended main chain of the peptide and helps to control the spacing between the binding subsites. The acidic-cluster-dileucine binding site is a distinctive motif for specific binding because it relies not on a single high-affinity interaction with one bulky side chain as a linchpin, but rather on a distributed set of several weaker interactions with medium-sized side chains. Specific recognition of acidic-cluster-dileucine signals by GGAs therefore seems to be a consequence of the rigid spacing of the three primary side-chain-binding subsites (Fig. 1f) together with the secondary subsites and the precise alignment of the peptide backbone to fulfil these interactions. □

Table 1 Affinity constants of peptides for GGA VHS domains

VHS domain construct	Peptide	K_d (μM)
GGA3	CI-MPR	10 ± 1
GGA3	CD-MPR	56 ± 11
GGA3	Amide-CI-MPR	>100*
GGA3	Amide-CD-MPR	>250*
GGA3-K130A	CI-MPR	>250*
GGA3-Y101A	CI-MPR	>250*
GGA3-N91A	CI-MPR	>250*
GGA1	CI-MPR	7.1 ± 0.8
GGA1	CD-MPR	29 ± 4
GGA1	Amide-CI-MPR	26 ± 4
GGA1	Amide-CD-MPR	>250*

* Affinity too weak for accurate measurement.

Methods

Protein expression, purification and crystallization

The VHS domain of GGA3 was cloned into the pHis-parallel2 vector[19]. The His_6-tagged VHS domain was expressed in BL21(DE3) cells at 21 °C and purified on a Ni^{2+}-nitrilotriacetate (NTA) column. The His_6 tag was removed with TEV protease and the protein was subjected to a second purification with Ni^{2+}-NTA. The expressed protein contains the vector-derived sequence GAMGS followed by GGA3 residues 1–166. Protein was dialysed into 50 mM NaCl, 20 mM Tris-HCl pH 8.0, 10 mM dithiothreitol (DTT). Peptide was added and the sample was diluted with dialysis buffer to 12 mg ml^{-1} protein and 1.5 mM peptide. VHS–peptide complexes crystallized in 2-μl hanging drops over a 0.5-ml reservoir of 100 mM CAPS (pH 10.2–11.0), 200 mM Li_2SO_4 and 1.33–2 M NaH_2PO_4/K_2HPO_4. The presence of peptides in the crystals was verified by electrospray mass spectrometry of crystals washed several times in mother liquor and dissolved in 5% acetic acid. Selenomethionylated protein was expressed in *Escherichia coli* B834(DE3) in defined media and purified and crystallized as above. Mutant GGA3-VHS domains were generated with the overlap extension technique or using the Quikchange site-directed mutagenesis kit (Stratagene) and verified by sequencing. Mutant VHS domains were expressed and purified in the same manner as the native domain.

X-ray data collection and structure solution

Crystals were cryoprotected in mother liquor supplemented with 20% glycerol and frozen directly in the cryostream. Native data were collected in-house by using CuKα radiation from a Rigaku RU-200 rotating-anode source and a Rigaku RAXIS-IV detector equipped with Osmic mirror optics. Multiwavelength anomalous dispersion (MAD) data were collected at beamline X9B, National Synchrotron Light Source. Data were collected at the selenium edge by using an ADSC Quantum 4 CCD detector. Two data sets were collected from crystals of the SeMet-VHS/CI-MPR peptide complex: one from a crystal belonging to space group $C222_1$ and a second that belonged to space group $P2_12_12$ (4 molecules per asymmetric unit in both space groups). All native and MAD data were indexed and reduced with HKL[20]. Phases were obtained from the $P2_12_12$ MAD dataset by using SOLVE (ref. 21, and http://www.solve.lanl.gov); solvent modification was performed with RESOLVE[22]. A model was built by using O[23] and refined by torsional dynamics and the maximum-likelihood target function with CNS[24]. Phases were subsequently obtained from the $C222_1$ MAD dataset by using SOLVE and RESOLVE, and the model from the other crystal was adjusted to fit the resulting solvent-modified map. The CD-MPR peptide was built and the model was refined against a laboratory CuKα $C222_1$ data set from a crystal of SeMet-VHS–CD-MPR peptide complex. The crystallographic R_{work} and R_{free} values for the VHS–CI-MPR complex are 21.4 and 25.8, respectively (35–2.4 Å data). The R_{work} and R_{free} values for the VHS–CD-MPR complex are 22.4 and 25.6, respectively (90–2.2 Å data). Other data-collection and refinement statistics are shown in the Supplementary Information. Molecular graphics in Figs 1 and 2 were rendered with SPOCK (http://quorum.tamu. edu/jon/spock), Bobscript[25] and Raster3D[26].

Yeast two-hybrid experiments

Yeast cells were co-transformed with plasmids encoding GAL4bd fused to the last 113 amino acids of the cytosolic tail of the CI-MPR, and GAL4ad fused to different mutants of the VHS domains of GGA1, GGA2 and GGA3. Cotransformed cells were spotted on plates lacking leucine and tryptophan, with or without histidine (+His and –His, respectively), then incubated at 30 °C. The construct GAL4bd-CI-MPR has been described previously[7]. The different GAL4ad-VHS constructs were generated with the use of the overlap extension technique and appropriate oligonucleotide primers. The cDNA fragments were cloned into the *Eco*R1/*Sal*I sites of the pGAD424 (LEU2) vector (Clontech, Palo Alto, CA).

Isothermal titration calorimetry

Wild-type and mutant GGA3 or wild-type GGA1 VHS domains were dialysed into 100 mM NaCl, 50 mM phosphate buffer (pH 7.5), 1 mM DTT to a concentration of 50 μM. Peptides were dissolved in the same buffer to 1 mM. Experiments were performed on a MicroCal MCS-ITC isothermal titration calorimeter at 30 °C and were analysed with Origin version 2.9 (MicroCal). Peptides were injected into 1.4 ml of the GGA VHS domains in 10-μl amounts for 21 injections. Data from peptide injections into buffer blanks were subtracted from experimental data before analysis.

Received 16 November 2001; accepted 10 January 2002.

1. Kirchhausen, T. Clathrin. *Annu. Rev. Biochem.* **69,** 699–727 (2000).
2. Dell'Angelica, E. C. *et al.* GGAs: a family of ADP ribosylation factor-binding proteins related to adaptors and associated with the Golgi complex. *J. Cell Biol.* **149,** 81–94 (2000).
3. Hirst, J. *et al.* A family of proteins with gamma-adaptin and VHS domains that facilitate trafficking between the trans-Golgi network and the vacuole/lysosome. *J. Cell Biol.* **149,** 67–80 (2000).
4. Boman, A. L., Zhang, C.-J., Zhu, X. & Kahn, R. A. A family of ADP-ribosylation factor effectors that can alter membrane transport through the trans-Golgi. *Mol. Biol. Cell* **11,** 1241–1255 (2000).
5. Poussu, A., Lohi, O. & Lehto, V.-P. Vear, a novel Golgi-associated protein with VHS and gamma-adaptin 'ear' domains. *J. Biol. Chem.* **275,** 7176–7183 (2000).
6. Lohi, O. & Lehto, V.-P. VHS domain marks a group of proteins involved in endocytosis and vesicular trafficking. *FEBS Lett.* **440,** 255–257 (1998).
7. Puertollano, R., Aguilar, R. C., Gorshkova, I., Crouch, R. J. & Bonifacino, J. S. Sorting of mannose 6-phosphate receptors mediated by the GGAs. *Science* **292,** 1712–1716 (2001).
8. Zhu, Y., Doray, B., Poussu, A., Lehto, V.-P. & Kornfeld, S. Binding of GGA2 to the lysosomal enzyme sorting motif of the mannose 6-phosphate receptor. *Science* **292,** 1716–1718 (2001).
9. Takatsu, H., Katoh, Y., Shiba, Y. & Nakayama, K. GGA proteins interact with acidic di-leucine sequences within the cytoplasmic domains of sorting receptors through their VHS domains. *J. Biol. Chem.* **276,** 28541–28545 (2001).
10. Nielsen, M. S. *et al.* The sortilin cytoplasmic tail conveys Golgi-endosome transport and binds the VHS domain of the GGA2 sorting protein. *EMBO J.* **20,** 2180–2190 (2001).
11. Mao, Y. *et al.* Crystal structure of the VHS and FYVE tandem domains of Hrs, a protein involved in membrane trafficking and signal transduction. *Cell* **100,** 447–456 (2000).
12. Misra, S., Beach, B. M. & Hurley, J. H. Structure of the VHS domain of human Tom1 (target of myb 1): insights into interactions with proteins and membranes. *Biochemistry* **39,** 11282–11290 (2000).
13. Chen, H. J., Yuan, J. & Lobel, P. Systematic mutational analysis of the cation-independent mannose 6-phosphate/insulin-like growth factor II receptor cytoplasmic domain. An acidic cluster containing a key aspartate is important for function in lysosomal enzyme sorting. *J. Biol. Chem.* **272,** 7003–7012 (1997).
14. Johnson, K. F. & Kornfeld, S. The cytoplasmic tail of the mannose 6-phosphate/insulin-like growth factor-II receptor has two signals for lysosomal enzyme sorting in the Golgi. *J. Cell. Biol.* **119,** 249–257 (1992).
15. Meresse, S. & Hoflack, B. J. Phosphorylation of the cation-independent mannose 6-phosphate receptor is closely associated with its exit from the trans-Golgi network. *J. Cell Biol.* **120,** 67–75 (1993).
16. Owen, D. J. & Evans, P. R. A structural explanation for the recognition of tyrosine-based endocytotic signals. *Science* **282,** 1327–1332 (1998).
17. Gatto, G. J. Jr, Geisbrecht, B. V., Gould, S. J. & Berg, J. M. Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5. *Nature Struct. Biol.* **7,** 1091–1095 (2000).
18. Scheufler, C. *et al.* Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70–Hsp90 multichaperone machine. *Cell* **101,** 199–210 (2000).
19. Sheffield, P., Garrard, S. & Derewenda, Z. Overcoming expression and purification problems of RhoGDI using a family of 'parallel' expression vectors. *Prot. Express. Purif.* **15,** 34–39 (1999).
20. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
21. Terwilliger, T. C. & Berendzen., J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55,** 849–861 (1999).
22. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56,** 965–972 (2000).
23. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjelgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
24. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
25. Esnouf, R. M. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D* **55,** 938–940 (1999).
26. Merritt, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277,** 505–524 (1997).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com).

Acknowledgements

We thank K. R. Rajashankar, U. Ramagopal, R. Trievel and E. Jones for assistance with data collection at beamline X9B, NSLS, D. E. Anderson for mass spectrometry, A. Peterkofsky for use of and assistance with the titration calorimeter, R. Ghirlando for analytical ultracentrifugation, X. Zhu for sequencing services, and H. Bernstein and P. Sun for comments on the manuscript.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.H.H. (e-mail: jh8e@nih.gov). Structural coordinates have been deposited in the Protein Data Bank under codes 1JPL and 1JUQ.

Structural basis for recognition of acidic-cluster dileucine sequence by GGA1

Tomoo Shiba*†‡, Hiroyuki Takatsu‡§, Terukazu Nogi*, Naohiro Matsugaki*, Masato Kawasaki*, Noriyuki Igarashi*, Mamoru Suzuki*, Ryuichi Kato*, Thomas Earnest‖, Kazuhisa Nakayama§ & Soichi Wakatsuki*

* *Photon Factory, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki 305-0801, Japan*
† *Foundation for Advancement of International Science (FAIS), Tsukuba, Ibaraki 305-0062, Japan*
§ *Institute of Biological Sciences and Gene Research Center, University of Tsukuba, Tsukuba, Ibaraki 305-8572, Japan*
‖ *Advanced Light Source, Berkeley, Berkeley Center for Structural Biology, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
‡ *These authors contributed equally to this work*

GGAs (Golgi-localizing, γ-adaptin ear homology domain, ARF-interacting proteins) are critical for the transport of soluble proteins from the *trans*-Golgi network (TGN) to endosomes/lysosomes by means of interactions with TGN-sorting receptors, ADP-ribosylation factor (ARF), and clathrin[1,2]. The amino-terminal VHS domains of GGAs form complexes with the cytoplasmic domains of sorting receptors by recognizing acidic-cluster dileucine (ACLL) sequences[1–6]. Here we report the X-ray structure of the GGA1 VHS domain alone, and in complex with the carboxy-terminal peptide of cation-independent mannose 6-phosphate receptor containing an ACLL sequence. The VHS domain forms